On an Iterative method for Direction of Arrival Estimation using Multiple Frequencies

Fredrik Andersson, Marcus Carlsson, Jean-Yves Tourneret, Senior Member, IEEE and Herwig Wendt, Member, IEEE,

I. PROBLEM DESCRIPTION

Assume that receivers are located along a linear array with spacing $\Delta$, i.e. at positions $x_j = j\Delta$, $j = -J, \ldots, J$. Assume that there are $P$ transmitting devices located sufficiently far away so that the a plane wave approximation of the transmitted signals can be used at the receiver array, and denote the transmitted signal by $s_p$. Let $\theta_p$ denote the angles made by the linear array and the plane waves, cf. Figure 1. Let

$$\tilde{s}_p(\omega) = \int_{-\infty}^{\infty} s_p(t)e^{-2\pi i t \omega} dt,$$

the Fourier transforms of $s_p$, respectively. Let $r_p(x, t)$ denote the signal from the transmitter $s_p$ recorded at receiver at location $x$. From the plane wave assumption on the incident waves it follows that

$$\tilde{r}_p(x, \omega) = e^{2\pi i x \cos(\theta_p)}\tilde{s}_0,0(\omega),$$

a relation that holds for each frequency $\omega$.

Let

$$r(x, t) = \sum_{p=1}^{P} r_p(x, t)$$

denote the actual measurements. For the Fourier transforms (with respect to the time variable) $\tilde{r}(x, \omega)$ it then holds that for each fixed frequency $\omega$ the function $\tilde{r}(x, \omega)$ is a sum of $P$ exponentials, i.e.

$$\tilde{r}(x, \omega) = \sum_{p=1}^{P} \tilde{r}_0,0(\omega)e^{2\pi i x \cos(\theta_p)} = \sum_{p=1}^{P} c_{p,\omega}e^{x\eta_p},$$

where

$$\eta_p = 2\pi i \cos(\theta_p),$$

and $c_{p,\omega} = \tilde{r}_0,0(\omega)$.

In practice, measurements of $r(x, t)$ may contain noise. The objective of the paper is to estimate the parameters $\eta_p$, and thus obtain the directions $\theta_p$. The way that we will do this is to construct operators (matrices) from the measurements, and enforce special structure on the operators that will allow us to make estimations of $\eta_p$.

The Kronecker theorem relates Hankel matrices of a fixed rank to sampled values of sums of a fixed number of exponential functions. A Hankel matrix $A$ has constant values on the anti-diagonals, i.e., $A(j, k) = A(j'+k')$ if $j + k = j' + k'$. It can thus be generated elementwise from a vector $f$ such that $A(j, k) = f_{j+k-1}$, $1 \leq j, k \leq N + 1$. We denote denote the generation of the Hankel matrix $A$ from a vector $f$ by $A = Hf$, and we denote the (adjoint) operation of summing a matrix $A$ over its anti-diagonals with $H^*A$, i.e.,

$$H^*A(n) = \sum_{l_1+l_2 = n}^{} A(l_1, l_2).$$

Note that Hankel matrix are complex symmetric, i.e. if $A(j, k) = (k, j)$ if $A$ is a Hankel matrix. Complex symmetric matrices can be factorized as

$$A = U\Sigma U^*,$$

where $\Sigma$ is a diagonal matrix with non-increasing elements $\sigma_j$ on the diagonal, and where the column vectors $u_j$ are the so-called con-eigenvalues which are orthonormal and satisfying $A_{u_j} = \sigma_j u_j u_j^*$.

Kronecker’s theorem [?, ?] states that if the Hankel matrix $A = Hf$ is of rank $P$ then, with the exception of degenerate cases, there exists $\{\zeta_p\}_{p=1}^{P}$ and $\{\epsilon_{p}\}_{p=1}^{P}$ in $\mathbb{C}$ such that $f$ is the sampled from the function

$$f(x) = \sum_{p=1}^{P} c_p e^{x\epsilon_p}.$$

(1)

The converse holds as well.

Moreover, it holds (still generically) that if $\sigma_i = 0$ then it holds for the polynomial

$$Q_u(\eta) = \sum_{k=0}^{N} u_k(k+1)\eta^k$$

generated from $u_j$, that $Q_{u_j}(\eta) = 0$. This can be seen from using (1):

$$0 = (Au)(j) = \sum_{k=1}^{N+1} A(j, k)u_k(k) = \sum_{k=1}^{N+1} \sum_{p=1}^{P} c_p e^{x\epsilon_p(j+k-1)}u_k(k) = \sum_{p=1}^{P} c_p e^{x\epsilon_p(j+k-1)}u_k(k) = \sum_{p=1}^{P} c_p e^{x\epsilon_p(j+k-1)}Q_u(\zeta_p),$$

which hold for $1 \leq j \leq N + 1$. As long a $P < N + 1$, this is an overdetermined system for the values of $Q_{u_j}(\zeta_p)$, and with the exception that the system is degenerated, it means that

Marcus Carlsson and Fredrik Andersson are with the Centre for Mathematical Sciences, Lund University, Lund, Sweden(fa@maths.lth.se, marcus.carlsson@math.lu.se).

J.-Y. Tourneret is with the Signal and Communication Group of the IRIT Laboratory, Toulouse University, France.

H. Wendt is with CNRS and with the Signal and Communication Group of the IRIT Laboratory, CNRS UMR 5505, Toulouse University, France(firstname.lastname@irit.fr).

Manuscript received Month XX, 2013; revised Month XX, 2013.
\( Q \omega_1(\zeta_p) = 0 \). In the above example, a unit sampling distance was used, but the argument holds for arbitrary fixed distance.

This means that it is possible to find the exponentials parameters \( \eta_p \) for a function of the form (1) by sampling it, finding a vector in the nullspace of the Hankel matrix generated from the samples, and finding the roots of the polynomial generated by this vector. The number of roots of \( Q \) is generated from the samples, and finding the roots of the linear (vandermonde) system. The coefficients associated with \( \zeta_p \) in (1) can be found by solving a linear system. The coefficients associated with the false roots should be zero (in the absence of noise), and thus the false roots can be discarded.

Another important (non-degenerate) property of non-full rank Hankel matrices is that the con-eigenvectors \( u_1, \ldots, u_P \) are also linear combinations of exponentials, i.e.,

\[
u_1(j) = \sum_{p=1}^{P} d_{1,p} e^{s \omega j}.
\]

A requirement for two vectors (sampled functions) \( f_1 \) and \( f_2 \) to be generated by the same set of \( P \) exponentials is thus that there exists orthonormal vectors \( u_1, \ldots, u_P \) such that if \( U_P = [u_1, \ldots, u_P] \), then

\[
\mathcal{F}U_P H f_1 P U_P = H f_1, \quad \mathcal{F}U_P H f_2 P U_P = H f_2,
\]

where \( \mathcal{F}U_P = U_P \mathcal{F} \) is the projection onto the span of \( U_P \).

The conditions above mean that \( H f_1 \) and \( H f_2 \) both have full rank \( P \) (because of the projection operator \( U_P \mathcal{F} \) and that both \( f_1 \) and \( f_2 \) are generated by the exponentials that generates \( u_1, \ldots, u_P \). Clearly, the same argument can be used to compare if \( f_1, \ldots, f_M \) are generated by the same set of exponentials.

In the noise free case, estimates of the exponential parameters can thus be immediately obtained for each fixed value of \( \omega \). In the presence of noise, one can consider rank \( P \) approximations of Hankel matrices for estimating the exponential parameters [1]. Related approaches include the well-known ESPRIT [1] and root-MUSIC [1] methods.

One optimization formulation for the estimation of exponential parameters is

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| f - g \|_2^2 \\
\text{subject to} & \quad \text{rank}(H g) = P.
\end{align*}
\]

Although there are some estimates for how good the estimates of exponential parameters that are obtained by instance ESPRIT [1], they are rather the results of an algorithm instead of the solution to an optimization problem, and it is therefore difficult to make statements for instance about optimality.

In [1] a method for obtaining an approximate solution to (3) is presented using the Alternate Direction Method of Multipliers (ADMM). The idea was further developed in [1], where also a method for solving optimization problems reminiscent of (3) where the available data is not equally spaced. More specifically, assume that we have knowledge of the data at points \( x = (x(j))_{j=-J}^{J} \).

To phrase the problem we need to make use of an auxiliary equally spaced grid and interpolation. If we can assume that the sampling at the equally spaced grid is sufficiently dense to represent the exponentials that we want to recover (essentially the Nyquist sampling rate, but with proper adjustments for it being on a finite interval []) prolates). We can construct a interpolation matrix \( K \) that maps samples at an equally spaced grid to samples at the unequally spaced grid \( x \).

Let \( x^{eq}(j) = \frac{J}{2}, -J \leq j \leq J \) and let

\[
K(j, k) = \varphi(2J(x^{eq}(j) - x(k))).
\]

The function \( \varphi \) is a compactly supported function that is reminiscent of a sinc-function. In the case where \( \varphi \) satisfies the Strang-Fix conditions, \( \hat{\varphi}(0) = 1, \hat{\varphi'}(0) = \hat{\varphi}^{(m)}(0) = 0, \) then \( K \) is capable of reproducing polynomials of orders \( \leq m \). Ideally, we would like \( \varphi \) to satisfy the partition of unity property

\[
\sum_{k} \varphi(x + k) = 1.
\]

One example of a function that satisfy the above property is Key’s interpolating function

\[
\varphi_{\text{Key}}(x) = \begin{cases} 
3/2|x|^3 - 5/2|x|^2 + 1 & \text{if } 0 \leq |x| \leq 1, \\
-1/2|x|^3 + 5/2|x|^2 - 4|x| + 2 & \text{if } 1 \leq |x| \leq 2, \\
0 & \text{if } |x| > 2.
\end{cases}
\]

The approximation order for \( \varphi_{\text{Key}} \) is three. An alternative function, where it is to control the number of lobes by a integer parameter \( a \), is the Lanczos function

\[
\varphi_{\text{Lanczos}}(x) = \begin{cases} 
sinc(x) \sin(x/a) & \text{if } 0 \leq |x| < a, \\
0 & \text{if } |x| \geq a.
\end{cases}
\]

For \( \varphi_{\text{Lanczos}} \) the partition of unity property is only approximate.

For instance, this mean that it can not be used to interpolate constant functions correctly. A simple (but not optimal) remedy that is commonly used to account for this deficiency, is to normalize the elements of the matrix \( K \) by their row sums. This trick also accounts for boundary problems where the finite support of \( \varphi \) falls outside the equally spaced nodes \( x^{eq} \). For an overview of cardinal interpolation, cf. [1].

A counterpart of (3) would then be

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| f - Kg \|_2^2 \\
\text{subject to} & \quad \text{rank}(H g) = P.
\end{align*}
\]

II. THE ADMM FORMULATION FOR DOA

Assume that there are measurements of \( r(x(j), \omega_m) \) available for \( m = 1, \ldots, Q \) where \( \omega_m < \omega_{m+1}, m = 1, \ldots, M-1 \).

For each frequency \( \omega_m \), let

\[
x_m(j) = \frac{j}{2J \Delta} \omega_m
\]

with \( J \) and \( \Delta \) as in the previous section. All the nodes \( x_m(j) \) are then contained in the unit length interval \([-1/2, 1/2]\). Let \( N > J, \) and let

\[
x^{eq}(n) = \frac{n}{2N}, -N \leq n \leq N,
\]

be an equally spaced lattice. Let \( K_m \) be an interpolation matrix that interpolates from the equally spaced data \( x^{eq} \) to the (scaled) points \( x_m \). For notation purposes, let

\[
r(j, m) = r(x(j), \omega_m),
\]
and let \( r_m \) denote the \( 2J + 1 \) vector obtained by fixing \( m \).

Assuming that the measured data originate from \( P \) sources, but is contaminated by noise. To reduce the influence of the noise we want to use all the measurements \( r(x(j), \omega_m), j = -J, \ldots, J, \ m = 1, \ldots, M \). The direction of arrival problem can then be formulated as

\[
\min_{g} \sum_{m=1}^{M} \| r_m - K_m g_m \|^2_2 \quad \text{subject to} \quad \text{rank}(H g_m) \leq P.
\]

The matrices \( K_m \) interpolates the function represented by the vector \( g_m \) on the equally spaced grid \( x^{eq} \) to the equally spaced grid \( x_m \). The objective function above thus describes the discrepancy between the interpolation between function sampled at \( x^{eq}(g_m) \) and the measured data \( r_m \).

We now try to reformulate the problem above into a form that is suitable to address using ADMM. Let \( R \) denote a rank penalizing function on sequences of matrices \( \{A_m\}_{m=1}^M \), for instance the rank limiting function

\[
R(\{A_m\}) = R_P(\{A_m\}) = \left\{ \begin{array}{ll} \|U_P\| = \|u_1, \ldots, u_P\| : A_m = U_m A_m U_m^H, \forall m, \\
0 \quad \text{otherwise} \end{array} \right.
\]

or the sum of nuclear norms

\[
R(\{A_m\}) = R_s(\{A_m\}) = \sum_{m=1}^{M} \|A_m\|_*.
\]

Above, \( \|A\|_* = \sum_{i} \sigma_i(A) \) denotes the nuclear norm, and \( \sigma_l(A) \) denotes the singular values of the matrix \( A \). The role of \( R \) is to enforce finite ranks on the matrices \( A_m \), which in combination with the Hankel constraint means that the generating vectors \( g_m \) are linear combination of a few \( P \) in the case of \( R_P \), and few in the case of \( R_s \) exponentials. The advantage of using the nuclear norm \( R = R_s \) is that the problem becomes convex. However, the solutions obtained will typically be biased in the way that the magnitude of the linear coefficients are underestimated. One alternative for avoiding that bias is to use the \( R_s \) for the frequency estimation, and correct for the amplitudes of the coefficients by a least squares fit once the frequencies are determined. The bias in the selection procedure is still not removed, see for instance [7]. In this paper, we will use \( R = R_P \).

\[
\min_{g} \ R_P(\{A_m\}) + \sum_{m=1}^{M} \| r_m - K_m g_m \|^2_2 \quad \text{subject to} \quad \text{rank}(H g_m) \leq P.
\]

The rank constraint of (5) is then contained in \( R_P(\{A_m\}) \) while the condition \( A_m(k,l) = r_m(k+l), \ -N \leq k,l \leq N \)

as a scalar product on complex valued matrices of size \((2N+1) \times (2N+1)\).

For the ADMM formulation we define (the augmented Lagrangian)

\[
L(\{A_m\}, \{g_m\}, \{\Lambda_m\}) = R_P(\{A_m\}) + \sum_{m=1}^{M} \| r_m - K_m g_m \|^2_2 + \sum_{m=1}^{M} \langle \Lambda_m, A_m - H g_m \rangle + \rho \| A_m - H g_m \|^2_2,
\]

where \( \|A\|^2_P = (A, A)_F \) denotes the Frobenius norm. The ADMM for (6) then reads for \( 1 \leq m \leq M \)

1) \( \{A_m^{k+1}\} = \arg\min_{A_m} L(\{A_m\}, \{g_m^k\}, \{\Lambda_m^k\}) \),

2) \( \{g_m^{k+1}\} = \arg\min_{g_m} L(\{A_m^{k+1}\}, \{g_m^k\}, \{\Lambda_m^k\}) \),

3) \( \Lambda_m^{k+1} = \Lambda_m^k + \rho (A_m^{k+1} - H g_m^{k+1}) \).

For the first step, by dropping the terms that are independent of \( A_m \) and by completion of squares we note that

\[
A_m^{k+1} = \arg\min_{A_m} R_P(\{A_m\}) + \sum_{m} \rho \bigg( A_m - \left( H g_m - \frac{\Lambda_m}{\rho} \right) \bigg)^2.
\]

The solution to this minimization step can be obtained by letting

\[
C = \sum_{m} \left( H g_m - \frac{\Lambda_m}{\rho} \right)^* \left( H g_m - \frac{\Lambda_m}{\rho} \right).
\]

Since \( C \) is a Hermitian matrix it can be diagonalized as

\[
C = U \Sigma U^* ,
\]

where \( \Sigma \) is a diagonal matrix containing the eigenvalues of \( C \) in descending order, and where the corresponding eigenvectors are the orthonormal columns of \( U \). Let \( U_P = [u_1, \ldots, u_P] \). The matrices \( A_m \) that solves (7) are then given by a projection of the complex symmetric matrix \( A_m^{k+1} = H g_m^{k+1} \) to Takagi factorizations using the span of the \( P \) eigenvectors of \( C \) corresponding to the \( P \) largest eigenvalues of \( C \), i.e.,

\[
A_m = \overline{U_P} \left( H g_m - \frac{\Lambda_m}{\rho} \right) U_P.
\]

Next, we turn our focus on to the second ADMM step. In a similar way as for the first step, we drop terms independent on \( g_m \) and complete squares. In this case, the vectors \( g_m \) can be analyzed individually for each fixed \( m \).

\[
g_m^{k+1} = \arg\min_{g_m} \| K_m g_m - r_m \|^2 + \rho \bigg( H g_m - A_m^{k+1} - \frac{\Lambda_m}{\rho} \bigg)^2.
\]

Then it is easy to see (following the same idea as in []) that (8) is minimized by solving

\[
(\rho \text{ diag}(w) + K_m \Lambda_m^k) g_m = \rho H^* A_m + \Lambda_m^{k+1} + K_m^* r_m,
\]

where \( w(n) = N + 1 - n, \ -N \leq n \leq N \) is a triangle weight.
function [g,eta]=hadmm_doa(rxm,g,P,J,M,rho,n_iter),
1:2:for m=1:M,K(m)=phi(log(exp(2*J+1,J));
1:3:K(m)=diag(sum(K{m}));K(m)=zeros(J+1);
1:4:for m=1:M,K(m)=sinv((rho*diag([1:J+1,J:J+1]));end;
1:5:for iter=1:n_iter,C=zeros(J+1);
1:6:1.2283 1.1341 1.0301 0.9112 0.8078 0.6638 0.1667
1:7:U
1:8:U=P=u(:,1:P)*u(:,1:P)';
1:9:a quasi-random generator, with
1:10:parameters
1:11:presentation, we normalize units so that
1:12:∆ = 1
1:13:31 receivers (M
1:14:representative case, we will works with
1:15:possible. A proper routine should also deals with convergence
1:16:the interpolation matrix on line 2. It is written to be as short as
1:17:the direction of arrival estimation problem:
1:18:procedure allows us to state a short and simple algorithm for
1:19:one single simulation. In this case, white noise has been added to
1:20:the measurement vectors \( \tilde{r}_m \) with a SNR of 0 dB.
1:21:Next, we estimate the parameters \( \eta_p \) for each fixed fre-
1:22:frequency \( \omega \) using ESPRIT. We then use the estimated param-
1:23:extrapolate values to be used as initial values of \( \hat{g}_m \) (the third
1:24:input parameter in the function of Table I). We then estimate
1:25:\( \eta_g \) for all frequencies \( \omega \) at the same time using the function
1:26:of Table I.
1:27:The obtained results are shown the Figure 2. The estimated
1:28:values of \( \eta_p \) for each fixed frequency \( \omega_m \) obtained by esprit
1:29:displayed by gray dots. The correct values of \( \eta_p \) are shown by
1:30:red lines, and the real part of the estimated values obtained by
1:31:the algorithm of Table I are shown by blue lines. The lines are
1:32:discontinued along the y-axis at \( M/2 \) for illustration purposes,
1:33:since they are more or less on top of each other. We note that
1:34:the obtained estimates when using the data for all the
1:35:estimations at the same time is substantially better than the
1:36:individual estimates obtained by ESPRIT.
1:37:Next, we study the estimation performance for several
1:38:simulations. We add white noise to the data using different
1:39:SNR, and for each SNR we run 100 simulations. To measure
1:40:the estimation performance, for each estimated value of \( \eta_p \) we
1:41:compute the distances to the all the true values of \( \eta_p \) and use
1:42:the smallest distance as a metric of how good the estimation
1:43:is. We then compute the standard deviation of the estimates.
1:44:For the ESPRIT estimates, we compute the standard deviation
1:45:over the of both the 100 simulations, and the 100 frequencies.
1:46:Note that as the correct values of \( \eta_p \) have an average distance
1:47:slightly larger than one, which gives somewhat of a limit of
1:48:how badly estimates that can be obtained even without any
1:49:data at all available. In Table II the results for the simulations
1:50:are shown. We see that the proposed method is consistently
1:51:outperforms using ESPRIT for each frequency individually,
1:52:and that it gives rather good results even for low SNR.

TABLE I: Main function (in MATLAB syntax) for the approximation and frequency estimation using \( P \) directions of arrival

![Fig. 1: Results for the direction of arrival problem using \( M=100 \) frequencies. The gray circles indicates the locations obtained by estimating the directions using ESPRIT on each \( r_m \) individually. The red lines indicate the true locations and the blue lines the estimates obtained by the algorithm in Table I.](image1)

Having solved the two minimization steps in the ADMM procedure allows us to state a short and simple algorithm for the direction of arrival estimation problem:

This routine is self contained except for the construction of the interpolation matrix on line 2. It is written to be as short as possible. A proper routine should also deals with convergence testing, and possibly to update the parameter \( \rho \) to increase the convergence rate, cf. []. The obtained estimates \( \eta \) needs to be rescaled with respect to \( \omega \) and \( \Delta \).

III. NUMERICAL SIMULATIONS

In this section we show some numerical results. As a representative case, we will work with \( M = 100 \) frequencies, 31 receivers \( (J = 16) \) and \( P = 8 \) sources. To simplify the presentation, we normalize units so that \( \Delta = 1 \), and use an affine frequency range \( \omega \) with \( \omega(1) = 0.5 \) and \( \omega(M) = 1 \). The parameters \( \eta_p \) are generated with zero imaginary part by a quasi-random generator, with \( |\eta_p| \leq \frac{2J}{P} \). The measured data \( \tilde{r}(x,\omega) \) is then generated from the exponential with \( \eta_p \) along with a pseudo-random coefficients for the individual frequencies \( \tilde{f}_p(x,\omega) \). First, we show the result obtained from

![Fig. 2: The points \( x_m \).](image2)

one single simulation. In this case, white noise has been added to the measurement vectors \( r_m \) with a SNR of 0 dB.

Next, we estimate the parameters \( \eta_p \) for each fixed frequency \( \omega \) using ESPRIT. We then use the estimated parameters to extrapolate values to be used as initial values of \( \hat{g}_m \) (the third input parameter in the function of Table I). We then estimate \( \eta_g \) for all frequencies \( \omega \) at the same time using the function of Table I.

The obtained results are shown the Figure 2. The estimated values of \( \eta_p \) for each fixed frequency \( \omega_m \) obtained by esprit is displayed by gray dots. The correct values of \( \eta_p \) are shown by red lines, and the real part of the estimated values obtained by the algorithm of Table I are shown by blue lines. The lines are discontinued along the y-axis at \( M/2 \) for illustration purposes, since they are more or less on top of each other. We note that the obtained estimates when using the data for all the frequencies at the same time is substantially better than the individual estimates obtained by ESPRIT.

Next, we study the estimation performance for several simulations. We add white noise to the data using different SNR, and for each SNR we run 100 simulations. To measure the estimation performance, for each estimated value of \( \eta_p \) we compute the distances to the all the true values of \( \eta_p \) and use the smallest distance as a metric of how good the estimation is. We then compute the standard deviation of the estimates. For the ESPRIT estimates, we compute the standard deviation over the of both the 100 simulations, and the 100 frequencies. Note that as the correct values of \( \eta_p \) have an average distance slightly larger than one, which gives somewhat of a limit of how badly estimates that can be obtained even without any data at all available. In Table II the results for the simulations are shown. We see that the proposed method is consistently outperforms using ESPRIT for each frequency individually, and that it gives rather good results even for low SNR.

TABLE II: Standard deviation in estimated directions

<table>
<thead>
<tr>
<th>Method</th>
<th>SNR (dB)</th>
<th>-6</th>
<th>-3</th>
<th>0</th>
<th>3</th>
<th>6</th>
<th>10</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>ESPRIT</td>
<td>0.3301</td>
<td>0.1838</td>
<td>0.0205</td>
<td>0.0144</td>
<td>0.0091</td>
<td>0.0054</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

IV. CONCLUSIONS

We have presented a method for direction of arrival estimation using wideband measurements using an ADMM method.
The method performs well and is easy to implement.